

日 本 国 特 許 庁
JAPAN PATENT OFFICE

03.06.2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2003年 6月10日

出 願 番 号
Application Number: 特願2003-165582
[ST. 10/C]: [JP2003-165582]

出 願 人
Applicant(s): 株式会社ケンウッド

REC'D 22 JUL 2004

WIPO

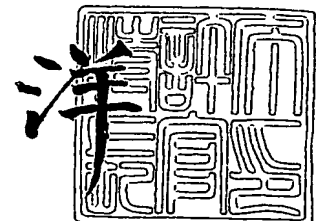
PCT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2004年 7月 9日

特許庁長官
Commissioner,
Japan Patent Office

小 川



【書類名】 特許願

【整理番号】 P07-975315

【提出日】 平成15年 6月10日

【あて先】 特許庁長官 殿

【国際特許分類】 G10L 13/06
G10L 13/08
G10L 11/04

【発明者】

【住所又は居所】 東京都八王子市石川町 2 9 6 7 - 3 株式会社ケンウッド内

【氏名】 佐藤 寧

【特許出願人】

【識別番号】 000003595

【氏名又は名称】 株式会社ケンウッド

【代理人】

【識別番号】 100095407

【弁理士】

【氏名又は名称】 木村 満

【手数料の表示】

【予納台帳番号】 038380

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9903184

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声選択装置、音声選択方法及びプログラム

【特許請求の範囲】

【請求項 1】

音声の波形を表す音声データを複数記憶する記憶手段と、
文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、
各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、
より構成されることを特徴とする音声選択装置。

【請求項 2】

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での 1 次回帰を行う回帰計算の結果に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものである、
ことを特徴とする請求項 1 に記載の音声選択装置。

【請求項 3】

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間の相関係数に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものである、
ことを特徴とする請求項 1 に記載の音声選択装置。

【請求項 4】

音声の波形を表す音声データを複数記憶する記憶手段と、
文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについ

ての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段と、より構成されており、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、ことを特徴とする音声選択装置。

【請求項 5】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での 1 次回帰により得られる 1 次関数の勾配からなる、

ことを特徴とする請求項 4 に記載の音声選択装置。

【請求項 6】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での 1 次回帰により得られる 1 次関数の切片からなる、

ことを特徴とする請求項 4 に記載の音声選択装置。

【請求項 7】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相関係数からなる、

ことを特徴とする請求項 4 に記載の音声選択装置。

【請求項 8】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデータを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最大値からなる、

ことを特徴とする請求項 4 に記載の音声選択装置。

【請求項 9】

前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶しており、

前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱う、

ことを特徴とする請求項 1 乃至 8 のいずれか 1 項に記載の音声選択装置。

【請求項 10】

選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備える、

ことを特徴とする請求項 1 乃至 9 のいずれか 1 項に記載の音声選択装置。

【請求項 11】

前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落部分合成手段を備え、

前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする請求項 10 に記載の音声選択装置。

【請求項 12】

音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測し、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する、

ことを特徴とする音声選択方法。

【請求項 13】

音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うこと

により、当該音片の時間長、及び、当該音片のピッチの時間変化を予測し、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する、ことを特徴とする音声選択方法であって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、ことを特徴とする音声選択方法。

【請求項 14】

コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、

して機能させるためのプログラム。

【請求項 15】

コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段として機能させるためのプログラムであって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値

の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、ことを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、音声選択装置、音声選択方法及びプログラムに関する。

【0002】

【従来の技術】

音声合成する手法として、録音編集方式と呼ばれる手法がある。録音編集方式は、駅の音声案内システムや、車載用のナビゲーション装置などに用いられている。

録音編集方式は、単語と、この単語を読み上げる音声を表す音声データとを対応付けておき、音声合成する対象の文章を単語に区切ってから、これらの単語に対応付けられた音声データを取得してつなぎ合わせる、という手法である（例えば、特許文献1参照）。

【0003】

【特許文献1】

特開平10-49193号公報

【0004】

【発明が解決しようとする課題】

しかし、音声データを単につなぎ合わせた場合、音声データ同士の境界では通常、音声のピッチ成分の周波数が不連続的に変化する、等の理由で、合成音声が不自然なものとなる。

この問題を解決する手法としては、同一の音素を互いに異なった韻律で読み上げる音声を表す複数の音声データを用意し、一方で音声合成する対象の文章に韻律予測を施して、予測結果に合致する音声データを選び出してつなぎ合わせる、という手法が考えられる。

しかし、音声データを音素毎に用意して録音編集方式により自然な合成音声を

得ようとする、音声データを記憶する記憶装置には膨大な記憶容量が必要となり、小型軽量の装置を用いる必要がある用途には適さない。また、検索する対象のデータの量も膨大なものとなるから、高速な処理が要求される用途にも適さない。

【0005】

この発明は、上記実状に鑑みてなされたものであり、簡単な構成で高速に自然な合成音声を得るための音声選択装置、音声選択方法及びプログラムを提供することを目的とする。

【0006】

【課題を解決するための手段】

上記目的を達成すべく、この発明の第1の観点にかかる音声選択装置は、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表して、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、より構成されることを特徴とする。

【0007】

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での1次回帰を行う回帰計算の結果に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものであってもよい。

【0008】

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間の相関係数に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものであってもよい。

【0009】

また、この発明の第2の観点にかかる音声選択装置は、
音声の波形を表す音声データを複数記憶する記憶手段と、
文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うこと
により、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測
手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについ
ての評価値を特定し、評価値が最も高い評価を表している音声データを選択する
選択手段と、より構成されており、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読み
が共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値
の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通す
る前記文章内の音片の時間長の予測結果との差の関数より得られるものである、
ことを特徴とする。

【0010】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音
片と読みが共通する前記文章内の音片のピッチの時間変化との間での1次回帰に
より得られる1次関数の勾配からなっているもよい。

【0011】

また、前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、
当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での1次
回帰により得られる1次関数の切片からなっているもよい。

【0012】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音
片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相
関係数からなっているもよい。

【0013】

前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデー
タを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通す
る前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最

大値からなっているもよい。

【0014】

前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶しているもよく、

前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱うものであってもよい。

【0015】

前記音声選択装置は、選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備えているもよい。

【0016】

前記音声選択装置は、前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落部分合成手段を備えているもよく、

前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成するものであってもよい。

【0017】

また、この発明の第3の観点にかかる音声選択方法は、
音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測し、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する、

ことを特徴とする。

【0018】

また、この発明の第4の観点にかかる音声選択方法は、

音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測し、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する、ことを特徴とする音声選択方法であって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、ことを特徴とする。

【0019】

また、この発明の第5の観点にかかるプログラムは、コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、

して機能させるためのものであることを特徴とする。

【0020】

また、この発明の第6の観点にかかるプログラムは、コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについ

ての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段として機能させるためのプログラムであって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、ことを特徴とする。

【 0 0 2 1 】

【発明の実施の形態】

以下、この発明の実施の形態を、音声合成システムを例とし、図面を参照して説明する。

図 1 は、この発明の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムは、本体ユニット M と、音片登録ユニット R とにより構成されている。

【 0 0 2 2 】

本体ユニット M は、言語処理部 1 と、一般単語辞書 2 と、ユーザ単語辞書 3 と、音響処理部 4 と、検索部 5 と、伸長部 6 と、波形データベース 7 と、音片編集部 8 と、検索部 9 と、音片データベース 1 0 と、話速変換部 1 1 とにより構成されている。

【 0 0 2 3 】

言語処理部 1、音響処理部 4、検索部 5、伸長部 6、音片編集部 8、検索部 9 及び話速変換部 1 1 は、いずれも、C P U (Central Processing Unit) や D S P (Digital Signal Processor) 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、言語処理部 1、音響処理部 4、検索部 5、伸長部 6、音片編集部 8、検索部 9 及び話速変換部 1 1 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。

【 0 0 2 4 】

一般単語辞書 2 は、P R O M (Programmable Read Only Memory) やハードディスク装置等の不揮発性メモリより構成されている。一般単語辞書 2 には、表意文字 (例えば、漢字など) を含む単語等と、この単語等の読みを表す表音文字 (例えば、カナや発音記号など) とが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。

【 0 0 2 5 】

ユーザ単語辞書 3 は、E E P R O M (Electrically Erasable/Programmable Read Only Memory) やハードディスク装置等のデータ書き換え可能な不揮発性メモリと、この不揮発性メモリへのデータの書き込みを制御する制御回路とにより構成されている。なお、プロセッサがこの制御回路の機能を行ってもよく、言語処理部 1、音響処理部 4、検索部 5、伸長部 6、音片編集部 8、検索部 9 及び話速変換部 1 1 の一部又は全部の機能を行うプロセッサがユーザ単語辞書 3 の制御回路の機能を行うようにしてもよい。

ユーザ単語辞書 3 は、表意文字を含む単語等と、この単語等の読みを表す表音文字とを、ユーザの操作に従って外部より取得し、互いに対応付けて記憶する。ユーザ単語辞書 3 には、一般単語辞書 2 に記憶されていない単語等とその読みを表す表音文字とが格納されていれば十分である。

【 0 0 2 6 】

波形データベース 7 は、P R O M やハードディスク装置等の不揮発性メモリより構成されている。波形データベース 7 には、表音文字と、この表音文字が表す単位音声の波形を表す波形データをエントロピー符号化して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。単位音声は、規則合成方式の手法で用いられる程度の短い音声であり、具体的には、音素や、V C V (Vowel-Consonant-Vowel) 音節などの単位で区切られる音声である。なお、エントロピー符号化される前の波形データは、例えば、P C M (Pulse Code Modulation) されたデジタル形式のデータからなっていればよい。

【 0 0 2 7 】

音片データベース 1 0 は、P R O M やハードディスク装置等の不揮発性メモリ

より構成されている。

音片データベース 10 には、例えば、図 2 に示すデータ構造を有するデータが記憶されている。すなわち、図示するように、音片データベース 10 に格納されているデータは、ヘッダ部 HDR、インデックス部 IDX、ディレクトリ部 DIR 及びデータ部 DAT の 4 種に分かれている。

【0028】

なお、音片データベース 10 へのデータの格納は、例えば、この音声合成システムの製造者によりあらかじめ行われ、及び／又は、音片登録ユニット R が後述する動作を行うことにより行われる。

【0029】

ヘッダ部 HDR には、音片データベース 10 を識別するデータや、インデックス部 IDX、ディレクトリ部 DIR 及びデータ部 DAT のデータ量、データの形式、著作権等の帰属などを示すデータが格納される。

【0030】

データ部 DAT には、音片の波形を表す音片データをエントロピー符号化して得られる圧縮音片データが格納されている。

なお、音片とは、音声のうち音素 1 個以上を含む連続した 1 区間をいい、通常は単語 1 個分又は複数個分の区間からなる。

また、エントロピー符号化される前の音片データは、上述の圧縮波形データの生成のためエントロピー符号化される前の波形データと同じ形式のデータ（例えば、PCM されたデジタル形式のデータ）からなっていればよい。

【0031】

ディレクトリ部 DIR には、個々の圧縮音声データについて、

- (A) この圧縮音片データが表す音片の読みを示す表音文字を表すデータ（音片読みデータ）、
- (B) この圧縮音片データが格納されている記憶位置の先頭のアドレスを表すデータ、
- (C) この圧縮音片データのデータ長を表すデータ、
- (D) この圧縮音片データが表す音片の発声スピード（再生した場合の時間長

)を表すデータ(スピード初期値データ)、

(E) この音片のピッチ成分の周波数の時間変化を表すデータ(ピッチ成分データ)、

が、互いに対応付けられた形で格納されている。(なお、音片データベース10の記憶領域にはアドレスが付されているものとする。)

【0032】

なお、図2は、データ部DATに含まれるデータとして、読みが「サイタマ」である音片の波形を表す、データ量1410hバイトの圧縮音片データが、アドレス001A36A6hを先頭とする論理的位置に格納されている場合を例示している。(なお、本明細書及び図面において、末尾に“h”を付した数字は16進数を表す。)

【0033】

また、ピッチ成分データは、例えば、図示するように、音片のピッチ成分の周波数をサンプリングして得られたサンプル $Y(i)$ (サンプルの総数を n として、 i は n 以下の正の整数)を表すデータであるものとする。

【0034】

なお、上述の(A)～(E)のデータの集合のうち少なくとも(A)のデータ(すなわち音片読みデータ)は、音片読みデータが表す表音文字に基づいて決められた順位に従ってソートされた状態で(例えば、表音文字がカナであれば、五十音順に従って、アドレス降順に並んだ状態で)、音片データベース10の記憶領域に格納されている。

【0035】

インデックス部IDXには、ディレトリ部DIRのデータのおおよその論理的位置を音片読みデータに基づいて特定するためのデータが格納されている。具体的には、例えば、音片読みデータがカナを表すものであるとして、カナ文字と、先頭1字がこのカナ文字であるような音片読みデータがどのような範囲のアドレスにあるかを示すデータとが、互いに対応付けて格納されている。

【0036】

なお、一般単語辞書2、ユーザ単語辞書3、波形データベース7及び音片デー

データベース 10 の一部又は全部の機能を単一の不揮発性メモリが行うようにしてもよい。

【0037】

音片データベース 10 へのデータの格納は、図 1 に示す音片登録ユニット R により行われる。音片登録ユニット R は、図示するように、収録音片データセット記憶部 12 と、音片データベース作成部 13 と、圧縮部 14 とにより構成されている。なお、音片登録ユニット R は音片データベース 10 とは着脱可能に接続されており、この場合は、音片データベース 10 に新たにデータを書き込むときを除いては、音片登録ユニット R を本体ユニット M から切り離した状態で本体ユニット M に後述の動作を行わせてよい。

【0038】

収録音片データセット記憶部 12 は、ハードディスク装置等のデータ書き換え可能な不揮発性メモリより構成されている。

収録音片データセット記憶部 12 には、音片の読みを表す表音文字と、この音片を人が実際に発声したものを集音して得た波形を表す音片データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、この音片データは、例えば、PCM されたデジタル形式のデータからなっていればよい。

【0039】

音片データベース作成部 13 及び圧縮部 14 は、CPU 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、このプログラムに従って後述する処理を行う。

【0040】

なお、音片データベース作成部 13 及び圧縮部 14 の一部又は全部の機能を単一のプロセッサが行うようにしてもよく、また、言語処理部 1、音響処理部 4、検索部 5、伸長部 6、音片編集部 8、検索部 9 及び話速変換部 11 の一部又は全部の機能を行うプロセッサが音片データベース作成部 13 や圧縮部 14 の機能を更に行ってもよい。また、音片データベース作成部 13 や圧縮部 14 の機能を行うプロセッサが、収録音片データセット記憶部 12 の制御回路の機能を兼ねても

よい。

【0041】

音片データベース作成部13は、収録音片データセット記憶部12より、互いに対応付けられている表音文字及び音片データを読み出し、この音片データが表す音声のピッチ成分の周波数の時間変化と、発声スピードとを特定する。

発声スピードの特定は、例えば、この音片データのサンプル数を数えることにより特定すればよい。

【0042】

一方、ピッチ成分の周波数の時間変化は、例えば、この音片データにケプストラム解析を施すことにより特定すればよい。具体的には、例えば、音片データが表す波形を時間軸上で多数の小部分へと区切り、得られたそれぞれの小部分の強度を、元の値の対数（対数の底は任意）に実質的に等しい値へと変換し、値が変換されたこの小部分のスペクトル（すなわち、ケプストラム）を、高速フーリエ変換の手法（あるいは、離散変数をフーリエ変換した結果を表すデータを生成する他の任意の手法）により求める。そして、このケプストラムの極大値を与える周波数のうちの最小値を、この小部分におけるピッチ成分の周波数として特定する。

【0043】

なお、ピッチ成分の周波数の時間変化は、例えば、特開2003-108172号公報に開示された手法に従って音片データをピッチ波形データへと変換してから、このピッチ波形データに基づいて特定するようにすると良好な結果が期待できる。具体的には、音片データをフィルタリングしてピッチ信号を抽出し、抽出されたピッチ信号に基づいて、音片データが表す波形を単位ピッチ長の区間へと区切り、各区間について、ピッチ信号との相関関係に基づいて位相のずれを特定して各区間の位相を揃えることにより、音片データをピッチ波形信号へと変換すればよい。そして、得られたピッチ波形信号を音片データとして扱い、ケプストラム解析を行う等することにより、ピッチ成分の周波数の時間変化を特定すればよい。

【0044】

一方、音片データベース作成部 13 は、収録音片データセット記憶部 12 より読み出した音片データを圧縮部 14 に供給する。

圧縮部 14 は、音片データベース作成部 13 より供給された音片データをエントロピー符号化して圧縮音片データを作成し、音片データベース作成部 13 に返送する。

【0045】

音片データの発声スピード及びピッチ成分の周波数の時間変化を特定し、この音片データがエントロピー符号化され圧縮音片データとなって圧縮部 14 より返送されると、音片データベース作成部 13 は、この圧縮音片データを、データ部 DAT を構成するデータとして、音片データベース 10 の記憶領域に書き込む。

【0046】

また、音片データベース作成部 13 は、書き込んだ圧縮音片データが表す音片の読みを示すものとして収録音片データセット記憶部 12 より読み出した表音文字を、音片読みデータとして音片データベース 10 の記憶領域に書き込む。

また、書き込んだ圧縮音片データの、音片データベース 10 の記憶領域内での先頭のアドレスを特定し、このアドレスを上述の (B) のデータとして音片データベース 10 の記憶領域に書き込む。

また、この圧縮音片データのデータ長を特定し、特定したデータ長を、(C) のデータとして音片データベース 10 の記憶領域に書き込む。

また、この圧縮音片データが表す音片の発声スピード及びピッチ成分の周波数の時間変化を特定した結果を示すデータを生成し、スピード初期値データ及びピッチ成分データとして音片データベース 10 の記憶領域に書き込む。

【0047】

次に、この音声合成システムの動作を説明する。

まず、言語処理部 1 が、この音声合成システムに音声を合成させる対象としてユーザが用意した、表意文字を含む文章（フリーテキスト）を記述したフリーテキストデータを外部から取得したとして説明する。

【0048】

なお、言語処理部 1 がフリーテキストデータを取得する手法は任意であり、例

例えば、図示しないインターフェース回路を介して外部の装置やネットワークから取得してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体（例えば、フロッピー（登録商標）ディスクやCD-ROMなど）から、この記録媒体ドライブ装置を介して読み取ってもよい。また、言語処理部1の機能を行っているプロセッサが、自ら実行している他の処理で用いたテキストデータを、フリーテキストデータとして、言語処理部1の処理へと引き渡すようにしてもよい。

【0049】

フリーテキストデータを取得すると、言語処理部1は、このフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定する。そして、この表意文字を、特定した表音文字へと置換する。そして、言語処理部1は、フリーテキスト内の表意文字がすべて表音文字へと置換した結果得られる表音文字列を、音響処理部4へと供給する。

【0050】

音響処理部4は、言語処理部1より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を検索するよう、検索部5に指示する。

【0051】

検索部5は、この指示に応答して波形データベース7を検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部6へと供給する。

【0052】

伸長部6は、検索部5より供給された圧縮波形データを、圧縮される前の波形データへと復元し、検索部5へと返送する。検索部5は、伸長部6より返送された波形データを、検索結果として音響処理部4へと供給する。

音響処理部4は、検索部5より供給された波形データを、言語処理部1より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部8へと供給する。

【0053】

音片編集部 8 は、音響処理部 4 より波形データを供給されると、この波形データを、供給された順序で互いに結合し、合成音声を表すデータ（合成音声データ）として出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

【0054】

なお、音片編集部 8 が合成音声データを出力する手法は任意であり、例えば、図示しない D/A (Digital-to-Analog) 変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、音片編集部 8 の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

【0055】

次に、音響処理部 4 が、外部より配信された、表音文字列を表すデータ（配信文字列データ）を取得したとする。（なお、音響処理部 4 が配信文字列データを取得する手法も任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で配信文字列データを取得すればよい。）

【0056】

この場合、音響処理部 4 は、配信文字列データが表す表音文字列を、言語処理部 1 より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字に対応する圧縮波形データが検索部 5 により索出され、圧縮される前の波形データが伸長部 6 により復元される。復元された各波形データは音響処理部 4 を介して音片編集部 8 へと供給され、音片編集部 8 が、この波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

【0057】

次に、音片編集部 8 が、定型メッセージデータ及び発声スピードデータを取得したとする。

なお、定型メッセージデータは、定型メッセージを表音文字列として表すデータであり、発声スピードデータは、定型メッセージデータが表す定型メッセージの発声スピードの指定値（この定型メッセージを発声する時間長の指定値）を示すデータである。

【0058】

また、音片編集部 8 が定型メッセージデータや発声スピードデータを取得する手法は任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータを取得すればよい。

【0059】

定型メッセージデータ及び発声スピードデータが音片編集部 8 に供給されると、音片編集部 8 は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部 9 に指示する。

【0060】

検索部 9 は、音片編集部 8 の指示に応答して音片データベース 10 を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとを索出し、索出された圧縮波形データを伸長部 6 へと供給する。1 個の音片につき複数の圧縮音片データが該当する場合も、該当する圧縮音片データすべてが、音声合成に用いられるデータの候補として索出される。一方、圧縮音片データを索出できなかった音片があった場合、検索部 9 は、該当する音片を識別するデータ（以下、欠落部分識別データと呼ぶ）を生成する。

【0061】

伸長部 6 は、検索部 9 より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部 9 へと返送する。検索部 9 は、伸長部 6 より返送され

た音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部 11 へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部 11 へと供給する。

【0062】

一方、音片編集部 8 は、話速変換部 11 に対し、話速変換部 11 に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

【0063】

話速変換部 11 は、音片編集部 8 の指示に応答し、検索部 9 より供給された音片データを指示に合致するように変換して、音片編集部 8 に供給する。具体的には、例えば、検索部 9 より供給された音片データの元の時間長を、索出されたスピード初期値データに基づいて特定した上、この音片データをリサンプリングして、この音片データのサンプル数を、音片編集部 8 の指示したスピードに合致する時間長にすればよい。

【0064】

また、話速変換部 11 は、検索部 9 より供給された音片読みデータ、スピード初期値データ及びピッチ成分データも音片編集部 8 に供給し、欠落部分識別データを検索部 9 より供給された場合は、更にこの欠落部分識別データも音片編集部 8 に供給する。

【0065】

なお、発声スピードデータが音片編集部 8 に供給されていない場合、音片編集部 8 は、話速変換部 11 に対し、話速変換部 11 に供給された音片データを変換せずに音片編集部 8 に供給するよう指示すればよく、話速変換部 11 は、この指示に応答し、検索部 9 より供給された音片データをそのまま音片編集部 8 に供給すればよい。

【0066】

音片編集部 8 は、話速変換部 11 より音片データ、音片読みデータ、スピード初期値データ及びピッチ成分データを供給されると、供給された音片データのう

ちから、定型メッセージを構成する音片の波形に最もよく近似できる波形を表す音片データを、音片 1 個につき 1 個ずつ選択する。

【0067】

具体的には、まず、音片編集部 8 は、定型メッセージデータが表す定型メッセージに、例えば「藤崎モデル」や「T o B I (Tone and Break Indices)」等の韻律予測の手法に基づいた解析を加えることにより、この定型メッセージ内の各音片のピッチ成分の周波数の時間変化を予測する。そして、音片毎に、ピッチ成分の周波数の時間変化の予測結果をサンプリングしたものを表すデジタル形式のデータ（以下、予測結果データと呼ぶ）を生成する。

【0068】

次に、音片編集部 8 は、定型メッセージ内のそれぞれの音片について、この音片のピッチ成分の周波数の時間変化の予測結果を表す予測結果データと、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化を表すピッチ成分データとの相関を求める。

【0069】

より具体的には、音片編集部 8 は、話速変換部 11 より供給された各々のピッチ成分データについて、例えば、数式 1 の右辺に示す値 α 及び数式 2 の右辺に示す値 β を求める。

【0070】

【数 1】

$$\alpha = \frac{\sum_{i=1}^n [\{X(i) - m_x\} \cdot \{Y(i) - m_y\}]}{\sum_{i=1}^n \{X(i) - m_x\}^2}$$

$$\text{ただし、} m_x = \sum_{i=1}^n \frac{X(i)}{n}, \quad m_y = \sum_{i=1}^n \frac{Y(i)}{n}$$

【0071】

【数 2】

$$\beta = m_y - (\alpha \cdot m_x)$$

【0072】

図3 (a) に示すように、ある音片についての予測結果データ（サンプルの総数は n 個とする）の i 番目のサンプルの値 $X(i)$ （ i は整数）の 1 次関数として、この音片と読みが合致する音片の波形を表す音片データについてのピッチ成分データ（サンプルの総数は n 個とする）の i 番目のサンプル $Y(i)$ の値を 1 次回帰させた場合、この 1 次関数の勾配は α 、切片は β となる。（勾配 α の単位は例えば [ヘルツ/秒] であればよく、切片 β の単位は例えば [ヘルツ] であればよい。）

【0073】

なお、同一の読みの音片について、予測結果データとピッチ成分データとでサンプルの総数が互いに異なる場合は、両者のうち一方（または両方）を、1 次補間やラグランジェ補間あるいはその他任意の手法により補間した上でリサンプリングし、両者のサンプルの総数を揃えてから相関を求めるようにすればよい。

【0074】

一方、音片編集部 8 は、話速変換部 11 より供給されたスピード初期値データ

と、音片編集部 8 に供給された定型メッセージデータ及び発声スピードデータとを用いて、数式 3 の右辺の値 d_t を求める。この値 d_t は、音片データが表す音片の発声スピードと、この音片と読みが合致する定型メッセージ内の音片の発声スピードとの時間差を表す係数である。

【0075】

【数 3】

$$d_t = |(X_t - Y_t) / Y_t|$$

(ただし、 Y_t は音片データが表す音片の発声スピード、 X_t はこの音片と読みが合致する定型メッセージ内の音片の発声スピード)

【0076】

そして、音片編集部 8 は、1 次回帰により得られた上述の α 及び β の値と、上述の係数 d_t とに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、数式 4 の右辺の値 (評価値) $cost_1$ が最大となるものを選択する。

【0077】

【数 4】

$$cost_1 = 1 / (W_1 |1 - \alpha| + W_2 |\beta| + d_t)$$

(ただし、 W_1 及び W_2 は所定の正の係数)

【0078】

音片のピッチ成分の周波数の時間変化の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化とが互いに近いほど、勾配 α の値は 1 に近くなり、従って、値 $|1 - \alpha|$ は 0 に近くなる。そして、評価値 $cost_1$ は、音片のピッチの予測結果と音片データのピッチとの相関が高いほど大きな値となるようにするため、値 $|1 - \alpha|$ の 1 次関数の逆数の形をとっているので、評価値 $cost_1$ は、値 $|1 - \alpha|$ が 0 に近くなるほど大きな値となる。

一方、音声の抑揚は、音片のピッチ成分の周波数の時間変化により特徴付けられる。従って、勾配 α の値は、音声の抑揚の差異を敏感に反映する性質を有する。

このため、合成されるべき音声について抑揚の正確さが重視される場合（例えば、電子メール等のテキストを読み上げる音声を合成する場合等）は、上述の係数 W_1 の値をなるべく大きくすることが望ましい。

【0079】

これに対し、音片のピッチ成分の基本周波数（ベースピッチ周波数）の予測結果と、この音片と読みが合致する音片の波形を表す音片データのベースピッチ周波数とが互いに近いほど、切片 β の値は0に近くなる。従って、切片 β の値は、音声のベースピッチ周波数の差異を敏感に反映する性質を有する。一方、評価値 cost1 は、値 $|\beta|$ の1次関数の逆数とみることもできる形をとっているの
で、評価値 cost1 は、値 $|\beta|$ が0に近くなるほど大きな値となる。

一方、音声のベースピッチ周波数は、音声の話者の声質を支配する要因であり、話者の性別による差異も顕著である。

このため、合成されるべき音声についてベースピッチ周波数の正確さが重視される場合（例えば、合成音声の話者の性別や声質を明確にする必要がある場合など）は、上述の係数 W_2 の値をなるべく大きくすることが望ましい。

【0080】

動作の説明に戻ると、音片編集部8は、定型メッセージ内の音片の波形に近い波形を表す音片データを選択する一方で、話速変換部11より欠落部分識別データも供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部4に供給し、この音片の波形を合成するよう指示する。

【0081】

指示を受けた音響処理部4は、音片編集部8より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表す圧縮波形データが検索部5により索出され、この圧縮波形データが伸長部6により元の波形データへと復元され、検索部5を介して音響処理部4へと供給される。音響処理部4は、この波形データを音片編集部8へと供給する。

【0082】

音片編集部 8 は、音響処理部 4 より波形データを返送されると、この波形データと、話速変換部 11 より供給された音片データのうち音片編集部 8 が特定したものとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

【0083】

なお、話速変換部 11 より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部 4 に波形の合成を指示することなく直ちに、音片編集部 8 が特定した音片データを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

【0084】

以上説明した、この音声合成システムでは、音素より大きな単位であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につなが合わせられ、定型メッセージを読み上げる音声合成される。音片データベース 10 の記憶容量は、音素毎に波形を記憶する場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追従できる。

【0085】

また、音片の波形の予測結果と音片データとの相関を複数の評価基準（例えば、1 次回帰させた場合の勾配や切片による評価と、音片の時間差による評価、など）で評価した場合は、これらの評価の結果に食い違いが生じる場合が多々あり得る。しかし、この音声合成システムでは、複数の評価基準で評価した結果が 1 個の評価値に基づいて総合され、適正な評価が行われる。

【0086】

なお、この音声合成システムの構成は上述のものに限られない。

例えば、波形データや音片データは PCM 形式のデータである必要はなく、データ形式は任意である。

また、波形データベース 7 や音片データベース 10 は波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース

7や音片データベース10が波形データや音片データをデータ圧縮されていない状態で記憶している場合、本体ユニットMは伸長部6を備えている必要はない。

【0087】

また、音片データベース作成部13は、図示しない記録媒体ドライブ装置にセットされた記録媒体から、この記録媒体ドライブ装置を介して、音片データベース10に追加する新たな圧縮音片データの材料となる音片データや表音文字列を読み取ってもよい。

また、音片登録ユニットRは、必ずしも収録音片データセット記憶部12を備えている必要はない。

【0088】

また、音片編集部8は、特定の音片の韻律を表す韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合は、この韻律登録データが表す韻律を、韻律予測の結果として扱うようにしてもよい。

また、音片編集部8は、過去の韻律予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

【0089】

また、音片編集部8は、上述の α 及び β の値を求める代わりに、話速変換部11より供給された各々のピッチ成分データについて、例えば、数式5の右辺に示す値 $R_{xy}(j)$ を、 j の値を0以上 n 未満の各整数として、合計 n 個求め、得られた $R_{xy}(0)$ から $R_{xy}(n-1)$ までの n 個の相関係数のうちの最大値を特定するようにしてもよい。

【0090】

【数5】

$$R_{xy}(j) = \frac{\sum_{i=1}^n [\{X(i) - m_x\} \cdot \{Y_j(i) - m_y\}]}{\sqrt{\sum_{i=1}^n \{X(i) - m_x\}^2} \sqrt{\sum_{i=1}^n \{Y(i) - m_y\}^2}}$$

【0091】

$R_{xy}(j)$ は、ある音片についての予測結果データ（サンプル総数 n 個。なお、数式 5 における $X(i)$ は数式 1 におけるものと同一である）と、この音片と読みが合致する音片の波形を表す音片データについてのピッチ成分データ（サンプルの総数 n 個）を一定の方向へ j 個循環シフトして得られたサンプルの列（なお、数式 5 において $Y_j(i)$ は、このサンプルの列の i 番目のサンプルの値である）との相関係数の値である。

【0092】

なお、図 3 (b) は、 $R_{xy}(0)$ 及び $R_{xy}(j)$ の値を求めるために用いる予測結果データ及びピッチ成分データの値の一例を示すグラフである。ただし、 $Y(p)$ の値（ただし、 p は 1 以上 n 以下の整数）は、循環シフトを行う前のピッチ成分データの p 番目のサンプルの値である。従って、例えば、音片データのサンプルが時刻の早い順に並んでおり、循環シフトが下位方向（つまり時刻が遅い方）へ行われるものとすれば、 $j < p$ の場合は $Y_j(p) = Y(p-j)$ であり、一方、 $1 \leq p \leq j$ の場合は $Y_j(p) = Y(n-j+p)$ である。

【0093】

そして、音片編集部 8 は、上述の $R_{xy}(j)$ の最大値と、上述の係数 d_t とに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、数式 6 の右辺の値（評価値） $cost_2$ が最大となるものを選択すればよい。

【0094】

【数 6】

$$cost_2 = 1 / (W_3 | R_{max} | + d_t)$$

（ただし、 W_3 は所定の係数、 R_{max} は $R_{xy}(0) \sim R_{xy}(n-1)$ のうちの最大値）

【0095】

なお、音片編集部 8 は、必ずしもピッチ成分データを種々循環シフトしたものについて上述の相関係数を求める必要はなく、例えば、 $R_{xy}(0)$ の値をそのまま相関係数の最大値として扱うようにしてもよい。

【0096】

また、評価値 $cost1$ や $cost2$ は、係数 d_t の項を含まなくてもよく、この場合、音片編集部 8 は、係数 d_t を求める必要がない。

あるいは、音片編集部 8 は、係数 d_t の値をそのまま評価値として用いてもよく、この場合、音片編集部は、勾配 α や、切片 β や、 $R_{xy}(j)$ の値を求める必要がない。

【0097】

また、ピッチ成分データは音片データが表す音片のピッチ長の時間変化を表すデータであってもよい。この場合、音片編集部 8 は、予測結果データとして、音片のピッチ長の時間変化の予測結果を表すデータを作成するものとし、この音片と読みが合致する音片の波形を表す音片データのピッチ長の時間変化を表すピッチ成分データとの相関を求めるようにすればよい。

【0098】

また、音片データベース作成部 13 は、マイクロフォン、増幅器、サンプリング回路、A/D (Analog-to-Digital) コンバータ及び PCM エンコーダなどを備えていてもよい。この場合、音片データベース作成部 13 は、収録音片データセット記憶部 12 より音片データを取得する代わりに、自己のマイクロフォンが集音した音声を表す音声信号を増幅し、サンプリングして A/D 変換した後、サンプリングされた音声信号に PCM 変調を施すことにより、音片データを作成してもよい。

【0099】

また、音片編集部 8 は、音響処理部 4 より返送された波形データを話速変換部 11 に供給することにより、当該波形データが表す波形の時間長を、発声スピードデータが示すスピードに合致させるようにしてもよい。

【0100】

また、音片編集部 8 は、例えば、言語処理部 1 と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音片の波形に最も近い波形を表す音片データを、定型メッセージに含まれる音片の波形に最も近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことに

よって選択して、音声の合成に用いてもよい。

この場合、音響処理部 4 は、音片編集部 8 が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部 5 に索出させなくてもよい。なお、音片編集部 8 は、音響処理部 4 が合成しなくてよい音片を音響処理部 4 に通知し、音響処理部 4 はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

【0101】

また、音片編集部 8 は、例えば、音響処理部 4 と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる音片の波形に最も近い波形を表す音片データを、定型メッセージに含まれる音片の波形に最も近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部 4 は、音片編集部 8 が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部 5 に索出させなくてもよい。

【0102】

以上、この発明の実施の形態を説明したが、この発明にかかる音声選択装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。

例えば、パーソナルコンピュータに上述の言語処理部 1、一般単語辞書 2、ユーザ単語辞書 3、音響処理部 4、検索部 5、伸長部 6、波形データベース 7、音片編集部 8、検索部 9、音片データベース 10 及び話速変換部 11 の動作を実行させるためのプログラムを格納した媒体（CD-ROM、MO、フロッピー（登録商標）ディスク等）から該プログラムをインストールすることにより、上述の処理を実行する本体ユニット M を構成することができる。

また、パーソナルコンピュータに上述の収録音片データセット記憶部 12、音片データベース作成部 13 及び圧縮部 14 の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、上述の処理を実行する音片登録ユニット R を構成することができる。

【0103】

そして、これらのプログラムを実行し本体ユニットMや音片登録ユニットRとして機能するパーソナルコンピュータが、図1の音声合成システムの動作に相当する処理として、図4～図6に示す処理を行うものとする。

図4は、このパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

図5は、このパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

図6は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【0104】

すなわち、まず、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（図4、ステップS101）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップS102）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

【0105】

そして、このパーソナルコンピュータは、フリーテキスト内の表意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース7より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップS103）。

【0106】

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップS104）、復元された波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する（ステップS105）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

【0107】

また、このパーソナルコンピュータが、外部より、上述の配信文字列データを任意の手法で取得すると（図5、ステップS201）、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース7より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップS202）。

【0108】

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップS203）、復元された波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップS105の処理と同様の処理により出力する（ステップS204）。

【0109】

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（図6、ステップS301）、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する（ステップS302）。

【0110】

また、ステップS302では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、1個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

【0111】

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の音片データへと復元する（ステップS303）。そして、復元された音片データを、上述の音片編集部8が行う処理と同様の処理により変換して、当該

音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる（ステップS304）。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

【0112】

次に、次に、このパーソナルコンピュータは、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の音片編集部8が行う処理と同様の処理を行うことにより、音片1個につき1個ずつ選択する（ステップS305～S308）。

【0113】

すなわち、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する（ステップS305）。そして、定型メッセージ内のそれぞれの音片について、この音片のピッチ成分の周波数の時間変化の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化を表すピッチ成分データとの相関を求める（ステップS306）。より具体的には、索出された各々のピッチ成分データについて、例えば、上述した勾配 α 及び切片 β の値を求める。

【0114】

一方で、このパーソナルコンピュータは、索出されたスピード初期値データと、外部より取得した定型メッセージデータ及び発声スピードデータとを用いて、上述の値 d_t を求める（ステップS307）。

【0115】

そして、このパーソナルコンピュータは、ステップS306で求めた α 、 β の値、及び、ステップS307で求めた d_t の値に基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、上述の評価値 $cost1$ が最大となるものを選択する（ステップS308）。

【0116】

なお、このパーソナルコンピュータは、ステップS306で、上述の α 及び β の値を求める代わりに、上述の $R_{xy}(j)$ の最大値を求めるようにしてもよい。

。この場合は、ステップS308で、 $R_{xy}(j)$ の最大値と、ステップS307で求めた係数 d_t とに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、上述の評価値 $cost_2$ が最大となるものを選択すればよい。

【0117】

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップS202～S203の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する(ステップS309)。

【0118】

そして、このパーソナルコンピュータは、復元した波形データと、ステップS308で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する(ステップS310)。

【0119】

なお、パーソナルコンピュータに本体ユニットMや音片登録ユニットRの機能を行わせるプログラムは、例えば、通信回線の掲示板(BBS)にアップロードし、これを通信回線を介して配信してもよく、また、これらのプログラムを表す信号により搬送波を変調し、得られた変調波を伝送し、この変調波を受信した装置が変調波を復調してこれらのプログラムを復元するようにしてもよい。

そして、これらのプログラムを起動し、OSの制御下に、他のアプリケーションプログラムと同様に実行することにより、上述の処理を実行することができる。

【0120】

なお、OSが処理の一部を分担する場合、あるいは、OSが本願発明の1つの構成要素の一部を構成するような場合には、記録媒体には、その部分を除いたプログラムを格納してもよい。この場合も、この発明では、その記録媒体には、コ

ンピュータが実行する各機能又はステップを実行するためのプログラムが格納されているものとする。

【0121】

【発明の効果】

以上説明したように、この発明によれば、簡単な構成で高速に自然な合成音声を得るための音声選択装置、音声選択方法及びプログラムが実現される。

【図面の簡単な説明】

【図1】

この発明の実施の形態に係る音声合成システムの構成を示すブロック図である。

【図2】

音片データベースのデータ構造を模式的に示す図である。

【図3】

(a) は、音片についてのピッチ成分の周波数の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化とを1次回帰させる処理を説明するためのグラフであり、(b) は、相関係数を求めるために用いる予測結果データピッチ成分データの値の一例を示すグラフである。

【図4】

この発明の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

【図5】

この発明の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

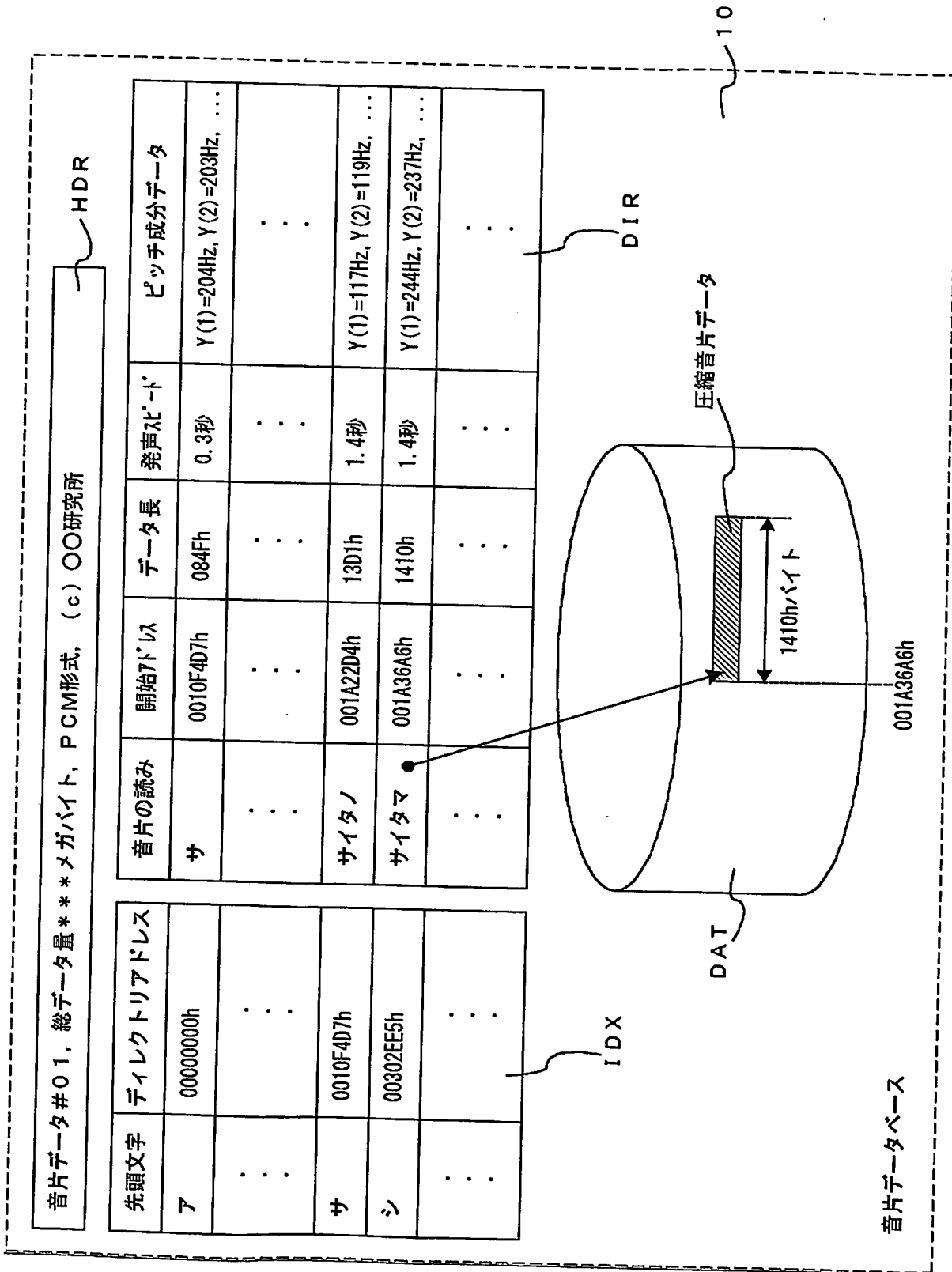
【図6】

この発明の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【符号の説明】

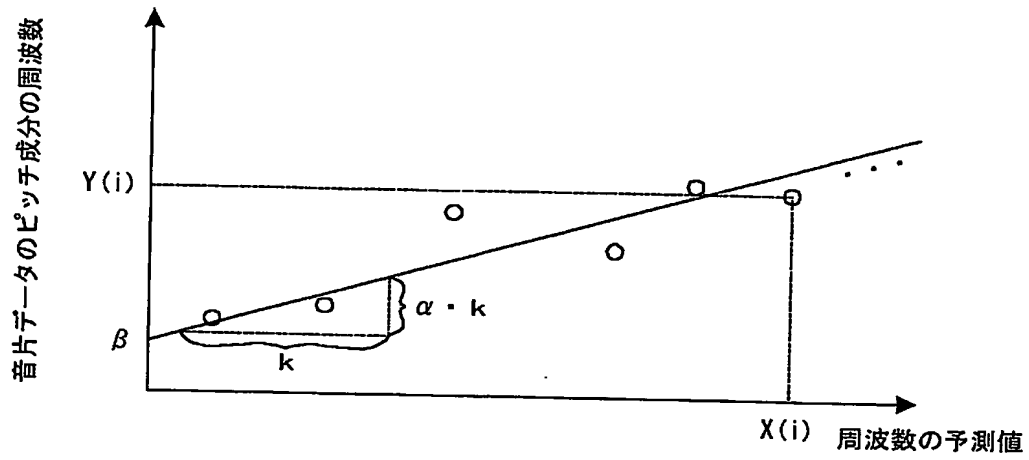
M	本体ユニット
1	言語処理部
2	一般単語辞書
3	ユーザ単語辞書
4	音響処理部
5	検索部
6	伸長部
7	波形データベース
8	音片編集部
9	検索部
1 0	音片データベース
1 1	話速変換部
R	音片登録ユニット
1 2	収録音片データセット記憶部
1 3	音片データベース作成部
1 4	圧縮部
H D R	ヘッダ部
I D X	インデックス部
D I R	ディレクトリ部
D A T	データ部

【図 2】

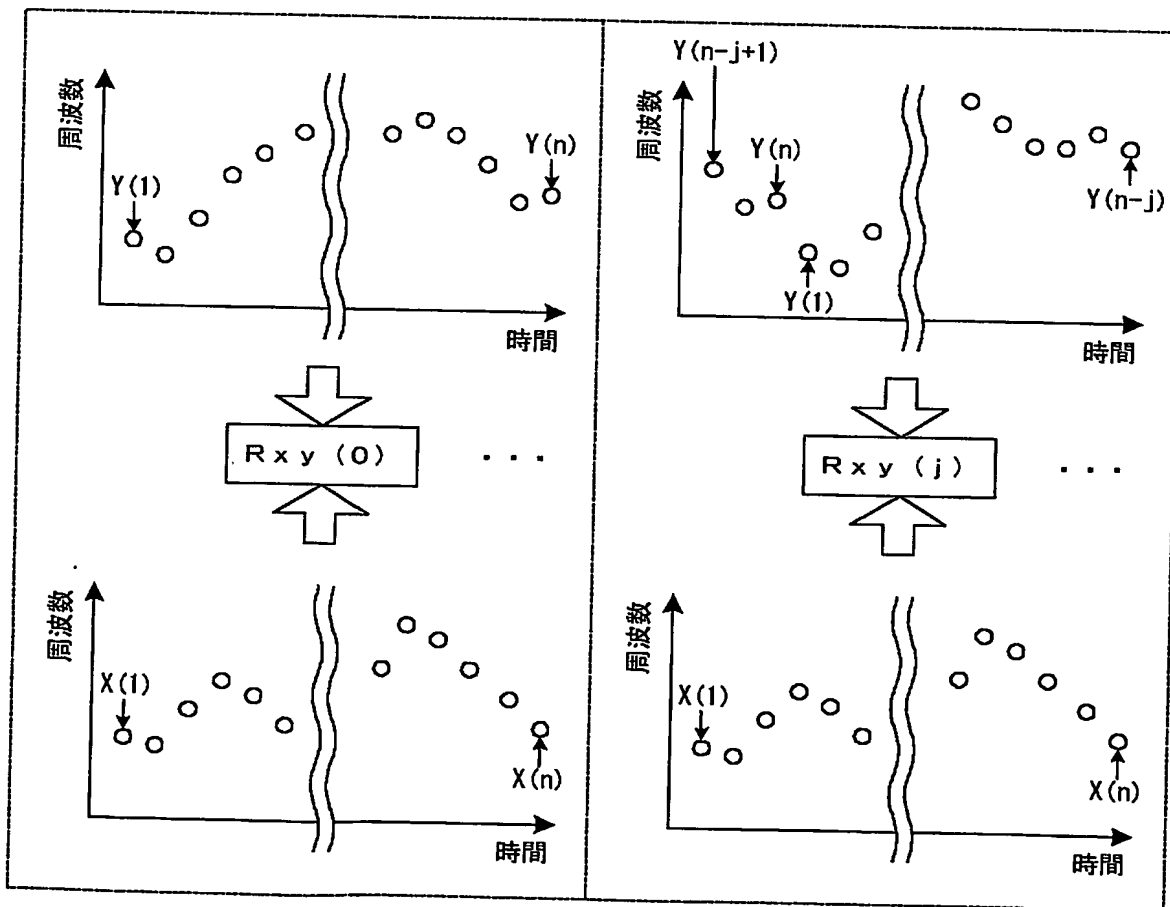


【図 3】

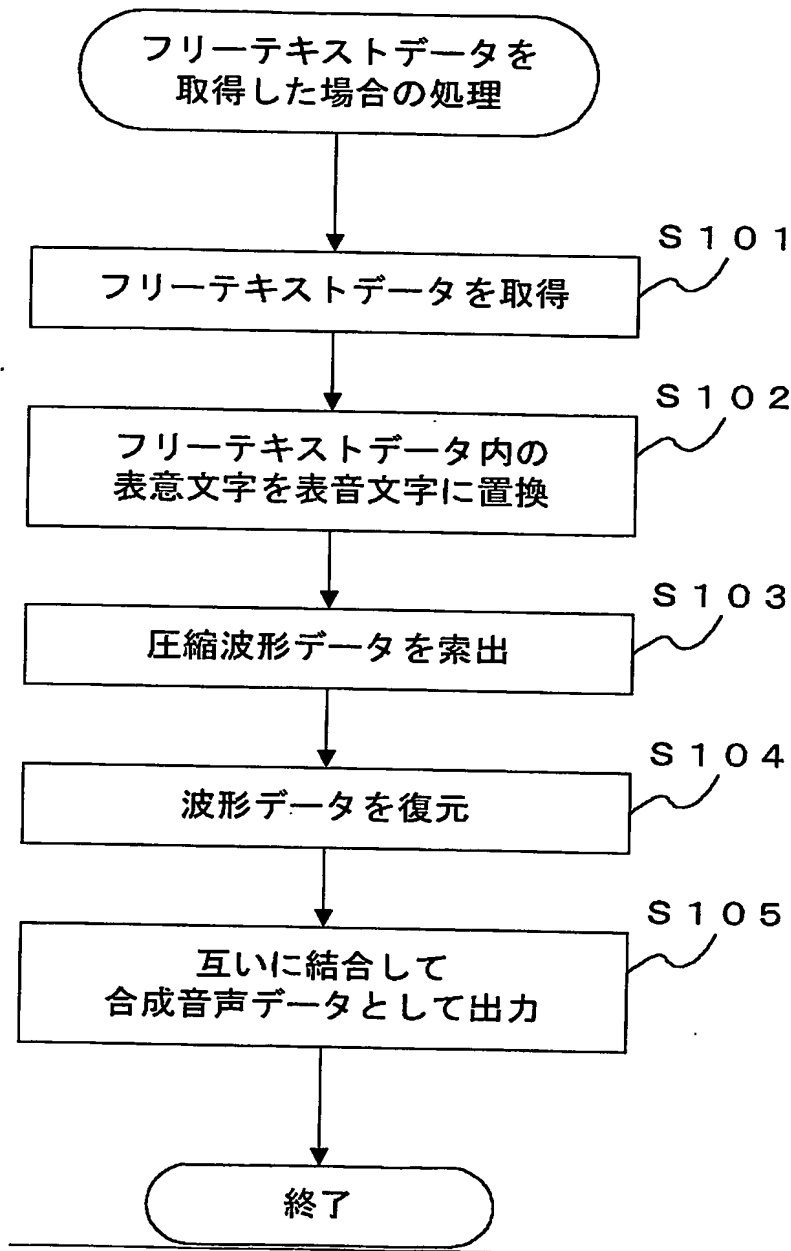
(a)



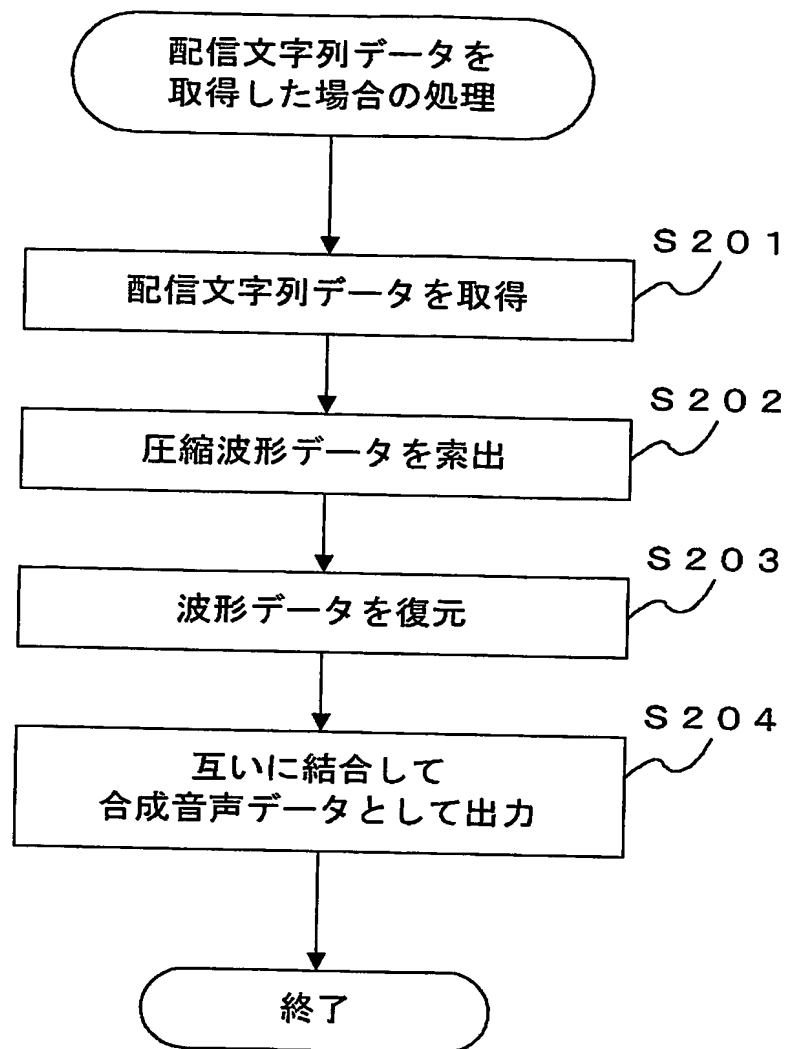
(b)



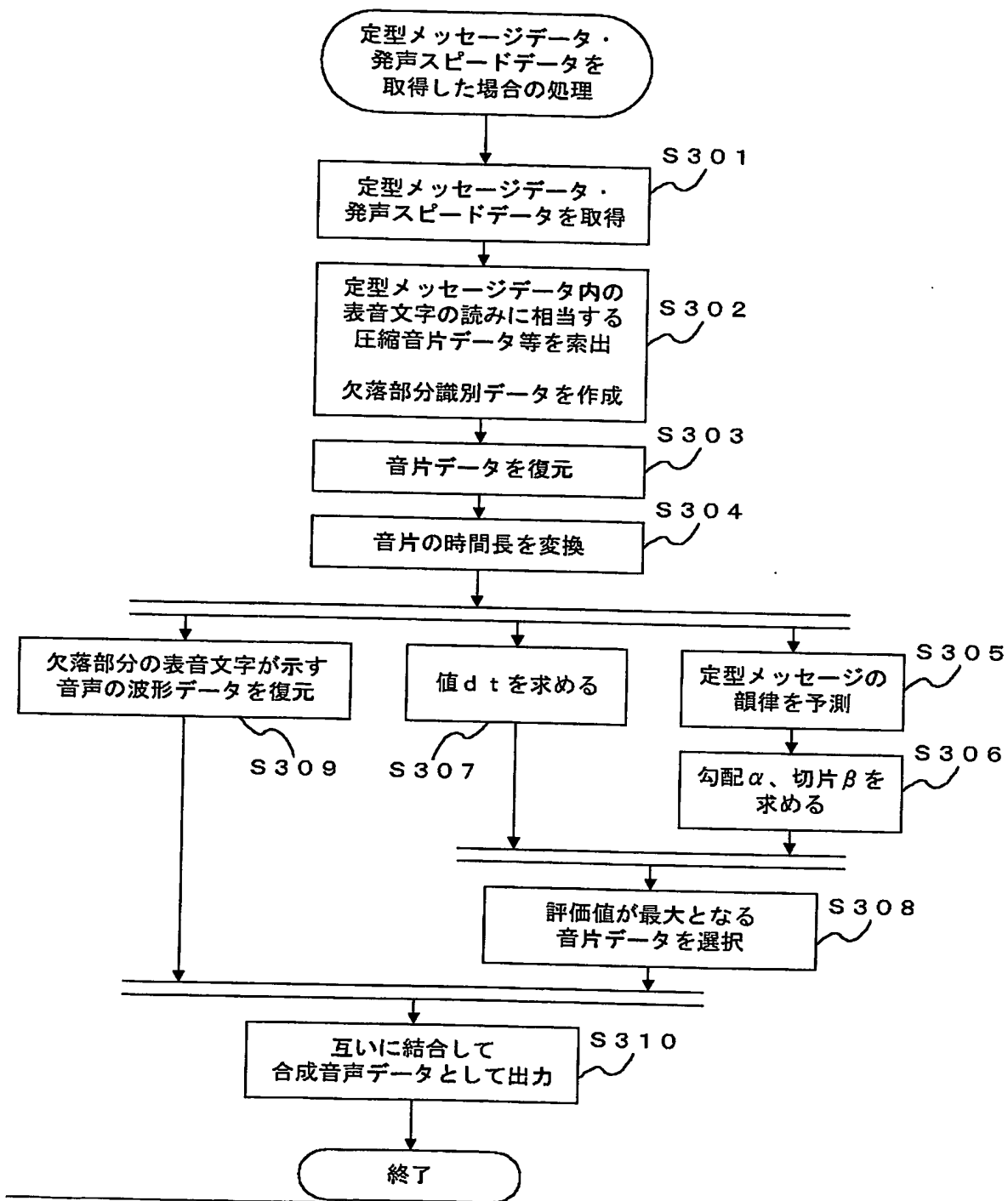
【図 4】



【図 5】



【図 6】



【書類名】 要約書

【要約】

【課題】 簡単な構成で高速に自然な合成音声を得るための音声選択装置等を提供することである。

【解決手段】 定型メッセージを表すデータが供給されると、音片編集部 8 は、定型メッセージ内の音片と読みが合致する音片の音片データを音片データベース 10 から索出させ、発声スピードデータが示すスピードに合致するよう、この音片データを変換させる。一方で音片編集部 8 は定型メッセージの韻律予測を行い、索出された音片データのうちから定型メッセージ内の各音片に最もよく合致するものを 1 個ずつ、評価式に基づいて特定する。評価式は、韻律予測結果－音片データ間におけるピッチ成分の周波数の 1 次回帰の結果や発声スピードの時間差を変数とするものである。そして、特定した音片データや、特定ができないため代わりに音響処理部 4 に供給させた波形データを互いに結合して、合成音声を表すデータを生成する。

【選択図】 図 1

特願 2003-165582

ページ : 1/E

出 願 人 履 歴 情 報

識別番号

[000003595]

1. 変更年月日

2002年 7月26日

[変更理由]

住所変更

住 所

東京都八王子市石川町2967番地3

氏 名

株式会社ケンウッド